Network Intrusion Detection with Minimal Communication Overhead

O. Patrick Kreidl and Alan S. Willsky

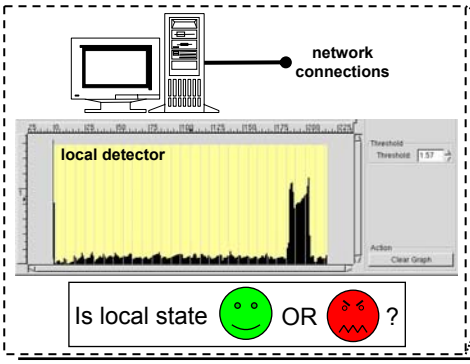
June 29, 2009

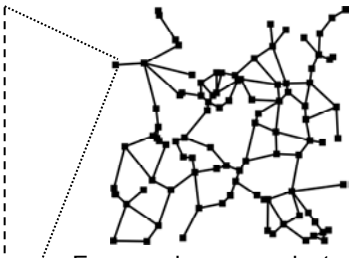
Workshop on Recent Advances in Intrusion Tolerance Systems,
IEEE Conference on Dependable Systems and Networks

Laboratory for Information and Decision Systems
Stochastic Systems Group

Problem Overview (distributed hypothesis test)

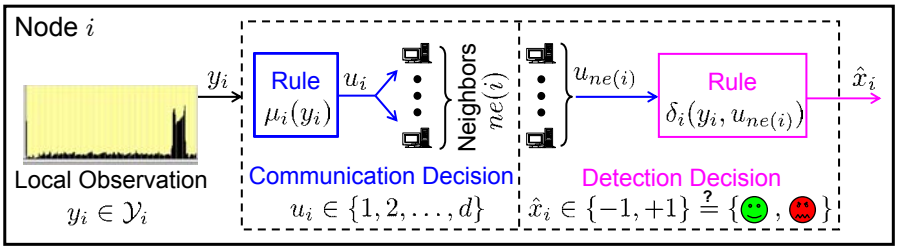
- Analyze a probabilistic model for network intrusion detection assuming
  - Each computer node hosts a local detector with tunable threshold
  - Constraints on allowable inter-node communication are explicit and severe





- Every node communicates only once & only to one-hop neighbors
- Every link is finite rate, $\log(d)$ bits

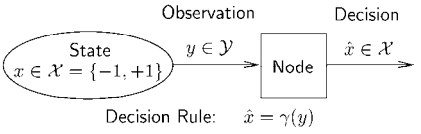
Solution Overview (under some assumptions)

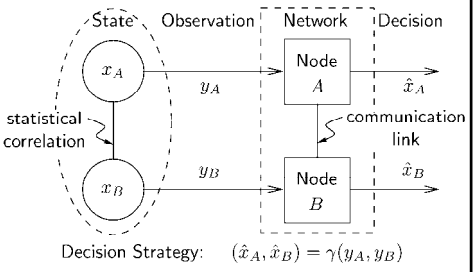


- Each node employs generalization of local *likelihood-ratio threshold rule*
  - Communication rule defined by up to d-1 distinct thresholds
  - Detection rule defined by up to $d^{|ne(i)|}$ distinct thresholds
- Team-optimal thresholds globally-coupled via system of nonlinear eqns
  - Depends on uncertain environment, network topology and error costs
  - Solved by iterative algorithm that is both efficient and convergent

Presentation Outline

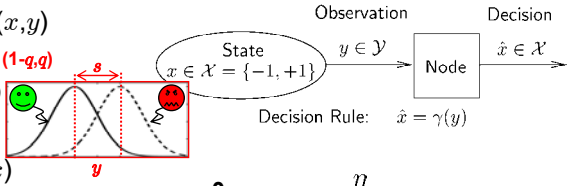
- Background: Bayesian Formulation
  - Probabilistic setup
  - Likelihood-ratio threshold rules
- Example: Two-Node Network
  - Problem parameterization
  - Performance analysis
- Closing Thoughts
  - Practical considerations
  - Implications for intrusion tolerance



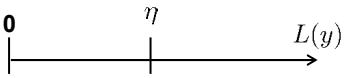


Background: Bayesian Formulation

- Given joint distribution $p(x,y)$
  - Prior probabilities $p(x)$ = (1-q,q)
  - Likelihood function $p(y|x)$



- Given cost function $c(\hat{x},x)$
  - Correct decisions are zero cost
  - False-negative cost ≥ false-positive cost
    = 1 = t



- Select rule $\gamma : \mathcal{Y} \to \mathcal{X}$ to minimize *risk* $J(\gamma) = \mathbf{E}\left[\mathbf{E}\left[c(\gamma(Y), X) | Y\right]\right]$

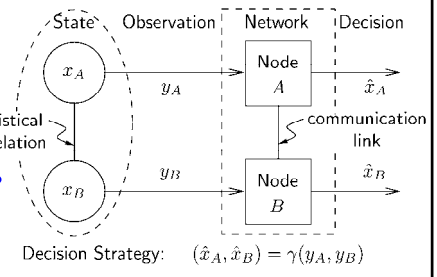
$$\exp(sy) = \underbrace{\frac{p_{Y|X}(y|+1)}{p_{Y|X}(y|-1)} = L(y)}_{\text{likelihood-ratio given } y} \quad \begin{matrix} \hat{x}=+1 \\ > \\ < \\ \hat{x}=-1 \end{matrix} \quad \underbrace{\eta = \frac{p_X(-1)c^{FP}}{p_X(+1)c^{FN}} = \left(\frac{1-q}{q}\right)t}_{\text{threshold}}$$

Two-Node Example: Problem Parameterization

- Assume $p(x,y) = p(x)p(y_A|x_A)p(y_B|x_B)$
  - Prior parameters q_A, q_B and r
  - Likelihood parameters s_A and s_B

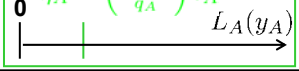
Locality/Separability Assumptions

- Assume $c(\hat{x},x) = c(\hat{x}_A,x_A) + c(\hat{x}_B,x_B)$
  - Correct decisions are zero cost
  - Error cost parameters t_A and t_B



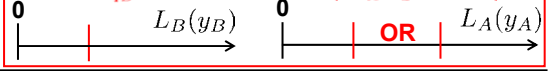
Myopic Strategy
(no communication)

$\eta_A = \left(\frac{1-q_A}{q_A}\right) t_A$



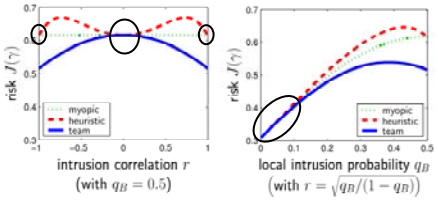
Heuristic Strategy
(send "local minimum-error" decision to the other)

$\theta_B = \frac{1-q_B}{q_B}$ $\quad \eta_A[u_B] = \left(\frac{p_{X_A,X_B}(-1,u_B)}{p_{X_A,X_B}(+1,u_B)}\right) t_A$

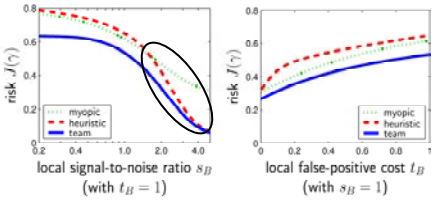


Two-Node Example: Performance Analysis

Risk vs. Prior Probabilities



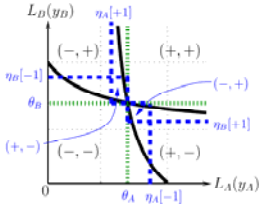
Risk vs. Likelihoods & Costs



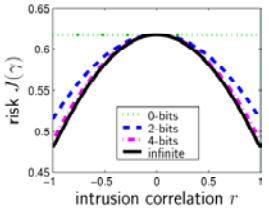
- Heuristic strategy performs worse than myopic strategy...except when
  - Two hidden states are weakly or maximally correlated
  - Observation at one node is much more informative than observation at other
- Team strategy always performs at least as well (and often notably better)

Two-Node Example: Insight into Team Strategy

Network-Constrained and Centralized Decision Regions



Team Performance vs. Communication Overhead



- Team strategy strives to best mimic centralized decision regions subject to allotted overhead (and also accounting for all uncertainty and costs)
- Team performance improves monotonically with additional overhead

Closing Thoughts

- Practical considerations
  - Locality/separability assumptions critical to scalable representation
    - Captures per-node compartmentalization of the observables/objectives
    - Implies no modeling challenges beyond the single-node case (almost)
  - Network sparsity is critical to efficiency of team optimization
    - Iterative algorithm scales exponentially in maximum node-degree
    - Initialization assumes each "neighborhood prior" $p(x_i, x_{ne(i)})$ is known
  - Typically different thresholds across all nodes (even if local IDSs are same)
- Implications for intrusion tolerance
  - Better intrusion detection will better cue tolerance-related controls
  - Can network attacks/detectors be adequately modeled probabilistically?
  - What is the analog "simplest model" for network intrusion tolerance?